

СТРУКТУРА МОДУЛЯ ИМПОРТА ДАННЫХ О ПУБЛИКАЦИОННОЙ АКТИВНОСТИ НАУЧНО-ПЕДАГОГИЧЕСКИХ РАБОТНИКОВ ИЗ РАЗЛИЧНЫХ НАУКОМЕТРИЧЕСКИХ СИСТЕМ

Аннотация. В настоящее время в рамках научных организаций, занимающихся научно-исследовательской деятельностью, активно используются наукометрические показатели публикационной активности. Под публикационной активностью обычно понимаются результаты научной или исследовательской деятельности автора, представленный в виде публикации. При формировании отчетной документации в рамках мониторинга деятельности образовательных организаций высшего образования проводится ручной сбор наукометрических показателей публикационной активности из реферативных и библиографических баз данных, находящихся в общем доступе в сети Интернет. Кроме этого, в определенных случаях однозначная идентификация причастности автора к организации, занимающийся научной деятельностью, становится затруднительной. Для решения этих проблем спроектирован и разработан модуль импорта из наукометрических систем, входящий в состав автоматизированной системы учета публикационной активности профессорско-преподавательского состава учебного заведения.

Ключевые слова: публикационная активность, наукометрические показатели, обработка данных, модуль импорта данных, наукометрические системы, Scopus, Web of Science, РИНЦ.

Abstract. Currently, within the framework of scientific organizations engaged in research activities, scientometric indicators of publication activity are actively used. Publication activity is usually understood as the results of scientific or research activities of the author, presented in the form of a publication. When generating reporting documentation within the framework of monitoring the activities of educational institutions of higher education, scientometric indicators of publication activity are manually collected from abstract and bibliographic databases that are publicly available on the Internet. In addition, in certain cases, the unambiguous identification of the author's involvement in an organization engaged in scientific activity becomes difficult. To solve these problems, a module for importing from scientometric systems was designed and developed, which is part of the automated system for recording the publication activity of the teaching staff of the educational institution.

Key words: publication activity, scientometric indicators, data processing, data import module, scientometric systems, Scopus, Web of Science, RSCI.

Введение. Одним из модулей автоматизированной системы мониторинга публикационной активности является подсистема импорта данных о наукометрических показателях. Подсистема обрабатывает информацию, сформированную в процессе экспорта из наукометрических систем, таких, как Scopus [1], Web of Science [2] и eLibrary [3]. Эти системы индексируют данные о публикационной активности из библиографических и реферативных баз данных, содержащих информацию о публикациях авторов со всего мира.

Поскольку эти системы могут индексировать разные публикации одного автора или содержать идентичные записи о публикациях, возникает проблема сбора наукометрических показателей. В первую очередь, это связано с тем, что необходимо учитывать публикации, которые индексируются сразу в нескольких библиографических базах данных [4]. Это вызывает проблему низкой достоверности информации о публикационной активности научно-педагогических работников. Главными причинами необходимости повышения достоверности можно обозначить:

1. Получение государственного задания от министерства науки и высшего образования;

2. Стимулирование научно-педагогических работников.

Существующие библиографические и реферативные базы данных, такие, как Scopus, Web of Science и РИНЦ [5], не позволяют в достаточной мере обеспечить повышение достоверности информации о публикационной активности. В первую очередь, это связано с проблемами сбора наукометрических показателей. Разработанное программное обеспечение позволяет объединить все данные из общедоступных библиографических и реферативных баз данных в рамках одной системы.

Особенности подсистемы импорта данных о публикационной активности. Подсистема импорта обрабатывает данные, сформированные в процессе экспорта из сторонних наукометрических систем. Данные представляют собой файл в формате, зависящем от системы. Так системы Web of Science и Scopus поддерживают импорт в формате BibTeX, а eLibrary в формате xml. Эти форматы были выбраны, как самые оптимальные для обработки в рамках одной системы. В качестве результата обработки выступают записи в базе данных системы, которая является отдельным модулем программного обеспечения.

Обработка представляет собой построчный анализ файлов, содержащих информацию о публикациях. Строки с данными содержат как названия показателей, так и их значения с разделителями. Для каждой публикации, вне зависимости от ее типа, можно выделить основные показатели, которые можно условно разделить на три категории: данные об авторах, основные показатели публикации и показатели издания. Так, каждая публикация, независимо от ее типа обладает показателями, представленными в виде ментальной карты на рисунке 1.

Процесс обработки заключается в разделении строки на две отдельные части: название показателя и его значение. Для каждой из наукометрических баз предусмотрен отдельный алгоритм обработки, поскольку каждая из систем формирует экспортные файлы в разном виде.

Таким образом, подсистема предназначена для обработки данных исключительно из заранее предусмотренных систем, таких как Scopus, Web of Science и eLibrary. В случаях, если необходимо обработать информацию, полученную из других источников, необходима модернизация программы и добавление новых методов обработки.

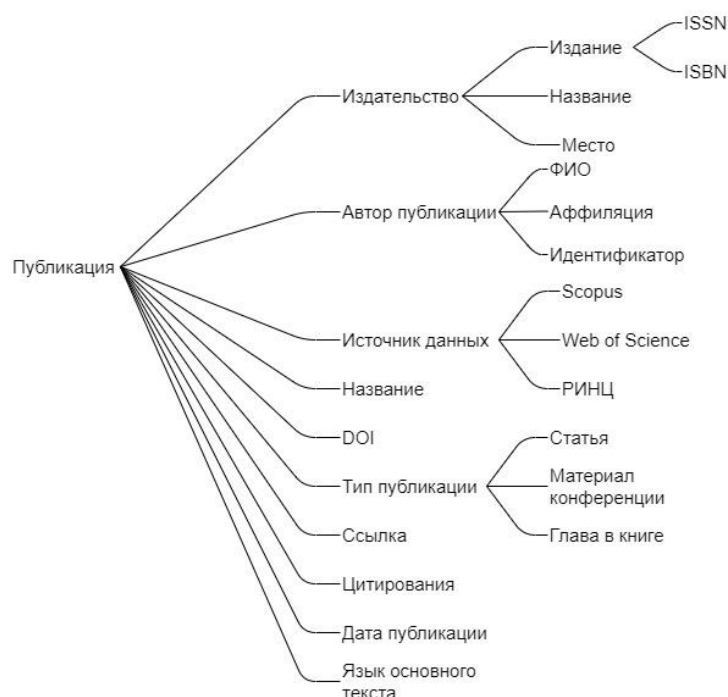


Рис. 1. Обобщенная древовидная структура информации о публикации

Схема работы подсистемы импорта данных. Для того, чтобы загрузить необходимую информацию в подсистему обработки и, как следствие, в базу данных пользователь должен предварительно зайти на сайт сторонней наукометрической системы и экспортировать оттуда данные о публикациях в соответствующем разделе сайта. Важно отметить, что пользователь должен загрузить данные исключительно в форматах, поддерживаемых подсистемой. После запуска модуля импорта пользователю предоставляется форма с полем для загрузки файла с предложением выбрать файл с компьютера. Полученные файлы пользователь добавляет на форму, после чего начинается процесс обработки. В случае появления исключений, представляющих собой дублирующие записи в базе данных или конфликты данных авторов, пользователю необходимо исправить их вручную.

Основной алгоритм обработки идентичен для каждой из сторонних наукометрических систем и представлен на рисунке 2. Основные различия в алгоритмах обработки обусловлены разным форматом хранения строк в файлах, сформированных в процессе экспорта из внешних систем.

В файлах в формате BibTeX, которые формируются в системах Scopus и Web of Science, данные разделены на блоки, которые представляют собой отдельную статью и разделены пустыми строками. Алгоритм обработки заключается в поиске пустых строк и посимвольном анализе блоков данных. Зачастую каждая строка в блоке представляет собой показатель и его значение с разделителями. Каждая строка анализируется посимвольно для разделения строки на показатель и его значение. Анализ строки разделяется на три этапа:

1. Поиск названия показателя.
2. Поиск открывающего и закрывающего разделителя.
3. Посимвольное добавление значения показателя в список.

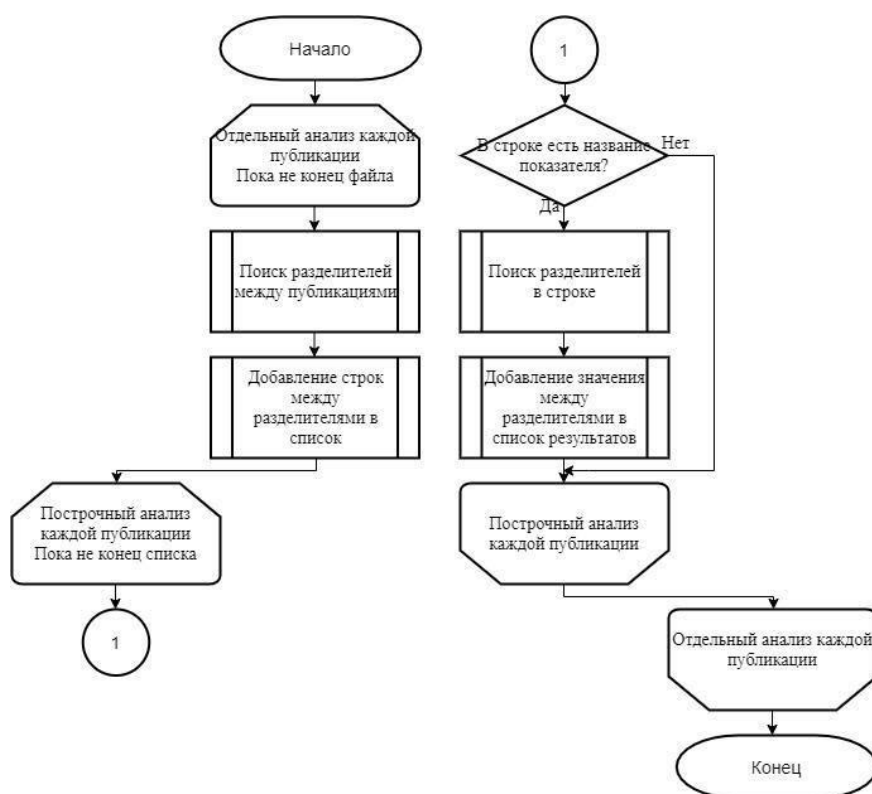


Рис. 2. Основной алгоритм обработки данных

Данные, прошедшие построчный анализ, необходимо обработать повторно. Это связано с тем, что некоторые показатели хранятся раздельно в базе данных системы. Так, фамилия, имя и отчество автора после первого этапа обработки хранятся в одной строке, в отличие от базы данных системы, которая предусматривает раздельное хранение этих данных. Аналогично хранятся и аффилиация, страницы публикации и информация о цитированиях. Еще одной целью второго этапа обработки является заполнение отсутствующих показателей пустыми значениями во избежание ошибок во время добавления показателей в базу данных.

Результаты. После всех этапов обработки данные проходят заключительный этап алгоритма, который заключается в добавлении данных из списков, в которых записывается обработанная информация, в базу данных системы. В соответствии со списком показателей, данные из списка значений этих показателей передаются в базу данных при помощи запросов. В качестве результата работы модуля импорта из наукометрических систем выступают непосредственно записи в базе данных автоматизированной системы.

База данных представляет собой отдельный модуль автоматизированной системы учета публикационной активности. После занесения записей в базу данных, они становятся доступными для работы. Так, например, к ним получает доступ подсистема составления отчетов, которая занимается расчетом показателей, необходимых для составления отчетной документации.

Обобщение. Таким образом, модуль импорта данных о публикационной активности собирает данные из сторонних систем и приводит их к единому формату. Обработанные данные поступают в остальные модули

автоматизированной системы, такие, как модуль базы данных и подсистема составления отчетов. Работа модуля существенно упрощает задачу подсистеме составления отчетов, поскольку все данные приведены к одному формату и хранятся в единой базе данных системы.

Список использованных источников

1. Data | Curated. Connected. Complete. – [Электронный ресурс]. – Режим доступа: <https://www.elsevier.com/solutions/scopus>. Дата обращения: 07.05.2021 г.
2. Добавление данных из системы Web of Knowledge / Web of Science. – [Электронный ресурс]. – Режим доступа: <https://istina.msu.ru/help/import/webofscience/>. Дата обращения: 07.05.2021 г.
3. Российский Индекс Научного Цитирования. – [Электронный ресурс]. – Режим доступа: https://www.elibrary.ru/project_risc.asp?. Дата обращения: 07.05.2021 г.
4. Основные положения при проектировании автоматизированной информационно-аналитической системы мониторинга и учета публикационной активности профессорско-преподавательского состава МГТУ им. Г.И. Носова / А.Б. Белявский, Н.С. Сибилева, И. Н. Орлова [и др.] // Новые информационные технологии и системы: сборник научных статей XVI Международной научно-технической конференции. – Пенза: Пензенский государственный университет, 2019. – С. 125-127.
5. Концепция разработки автоматизированной информационно-аналитической системы учета наукометрических показателей профессорско-преподавательского состава МГТУ им. Г.И. Носова / А.Б. Белявский, Н.С. Сибилева, И.Н. Орлова [и др.] // Ab ovo ... (С самого начала...): Сборник научных трудов. – Магнитогорск: Магнитогорский государственный технический университет им. Г.И. Носова, 2019. – С. 96-97.

УДК 504.064.36

Р. М. Эрлихман, В. Ю. Носков

ФГАОУ ВО «Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина», г. Екатеринбург, Россия

РАЗРАБОТКА СИСТЕМЫ МОНИТОРИНГА ТЕМПЕРАТУРЫ В ПОМЕЩЕНИЯХ С ИСПОЛЬЗОВАНИЕМ MESH-СЕТИ

Аннотация. *Определены основные задачи системы мониторинга температуры в помещениях с использованием mesh-сети. Рассмотрены используемые для решения этих задач инструменты. Представлена реализация и описание решения, позволяющего собирать и хранить данные о температуре в помещениях*

Ключевые слова: веб-сервис, язык программирования, Arduino IDE, микроконтроллер, датчик, mesh-сеть, ячеистая топология, MySQL.